NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

AN ANALYSIS OF ANAM READINESS EVALUATION SYSTEM (ARES) AS A PREDICTOR OF PERFORMANCE DEGRADATION INDUCED BY SLEEP DEPRIVATION IN OFFICER INDOCTRINATION SCHOOL (OIS) STUDENTS

by

Shonee L. Kenyon Yonkers

June 2004

Thesis Advisor:	Laura A. Barton
Second Reader:	Nita Lewis Miller

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** June 2004	**3. REPORT TYPE AND DATES COVERED** Master's Thesis

4. TITLE AND SUBTITLE: An Analysis of ANAM Readiness Evaluation System (ARES) as a Predictor of Performance Degradation Induced by Sleep Deprivation in Officer Indoctrination School (OIS) Students

5. FUNDING NUMBERS

6. AUTHOR(S) Shonee L. Kenyon Yonkers

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES)
N/A

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.	**12b. DISTRIBUTION CODE**

13. ABSTRACT (maximum 200 words)

Modeling fatigue, sleepiness, and performance is of significant interest to military leaders because military operations often provide limited sleep opportunities for many individuals. The ANAM Readiness Evaluation System (ARES) Commander Battery is under consideration as a quick, inexpensive method of testing a crewmember's level of functioning. This thesis analyzed data collected during a previous field fatigue study conducted at the Naval Officer Indoctrination School (OIS) in Newport, Rhode Island. Linear mixed-effects models were developed and ARES data were evaluated for how they vary across participants, testing sessions, and time of day.

14. SUBJECT TERMS
ARES, FAST, SAFTE, Alertness, ANAM, Sleep Regulation, Sleep Deprivation, Sleep Inertia, Sleep Debt, Performance, Linear Mixed Effects Model

15. NUMBER OF PAGES 75

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

AN ANALYSIS OF ANAM READINESS EVALUATION SYSTEM (ARES) AS A PREDICTOR OF PERFORMANCE DEGRADATION INDUCED BY SLEEP DEPRIVATION IN OFFICER INDOCTRINATION SCHOOL (OIS) STUDENTS

Shonee L. Kenyon Yonkers
Lieutenant, United States Navy Reserve
B.A., University of California San Diego, 1995
M.P.H, San Diego State University, 1997

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
June 2004

Author: Shonee L. Kenyon Yonkers

Approved by: Laura A. Barton
Thesis Advisor

Nita Lewis Miller
Second Reader

James N. Eagle
Chairman, Department of
Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Modeling fatigue, sleepiness, and performance is of significant interest to military leaders because military operations often provide limited sleep opportunities for many individuals. The ANAM Readiness Evaluation System (ARES) Commander Battery is under consideration as a quick, inexpensive method of testing a crewmember's level of functioning. This thesis analyzed data collected during a previous field fatigue study conducted at the Naval Officer Indoctrination School (OIS) in Newport, Rhode Island. Linear mixed-effects models were developed and ARES data were evaluated for how they vary across participants, testing sessions, and time of day.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

Modeling fatigue, sleepiness, and performance is of significant interest in the military operational community. Because a person is not a reliable judge of his or her own level of biological sleepiness, commanders require an objective means to assess their crewmembers' ability to perform. One such method is FAST, the software application based upon SAFTE™. SAFTE™ is a biomathematical model designed to predict individual and group performance under conditions of sleep deprivation. Also, psychomotor vigilance tests, such as the ARES Commander Battery, provide instant feedback on an individual's ability to sustain levels of concentration, working memory, and mental efficiency.

FAST is currently the preferred tool used to predict performance. However, days of sleep and activity data must be collected before a meaningful assessment can be produced. In contrast, the ARES Commander Battery takes less than 10 minutes and can be administered on a digital personal assistant. ARES is a new software package that has not been validated, but is under consideration as a quick, inexpensive method of testing an individual's level of functioning in a military operational setting.

Sleep and performance measures were collected during a previous study conducted in 2003 at Officer Indoctrination School (OIS) in Newport, Rhode Island. This thesis includes an analysis of the OIS data. Research goals consist of identifying how ARES Simple Reaction Time and Continuous Running Memory test scores vary by subject,

session, and time of day. Additionally, the relationship between ARES data and FAST performance effectiveness scores were explored. Mixed-effects modeling was employed in order to isolate variability due to both inter- and intra-individual differences.

Overall, the ARES variables, mean, median, and standard deviation of participants' reaction time for correct responses, show promise as instantaneous indicators of human performance decrement under conditions of mild sleep deprivation (i.e., an average of six hours per night). Also, it was discovered that throughput did not account for variance in FAST performance effectiveness. Finally, inter-individual differences accounted for a significant portion of the variability in ARES simple reaction time scores, but the session explained much of the variability in ARES continuous running memory scores, suggesting a possible learning effect.

ACKNOWLEDGMENTS

MAJ Maureen O'Connor and LT Paul Patillo designed and conducted the Navy Officer Indoctrination School (OIS) fatigue study. The data they collected made this thesis possible. Thank you for your data files and numerous references.

I am also appreciative of those who helped MAJ O'Connor and LT Patillo; I also have benefited from their support. Dr. Tim Elsmore, CDR Dennis Reeves, and Dr. Steve Hursh, I thank all of you. Additionally, LCDR Michael Lowe was generous with his time and ideas. He gave me an introduction to ANAM and ARES and provided an initial analysis of the OIS data.

Special thanks go to my witty Statistics professor, Dr. Sam Buttrey. Over the past year he's granted me dozens of unscheduled visits to discuss my data. I don't know that I'd have been able to navigate through SPLUS without him.

Thank you Dr. Nita Lewis Miller and LCDR Laura Barton for helping me pull it all together. Lastly, I thank my mom, grandma and Gus for entertaining the girls so I could work on my thesis. Savannah and Sasha, mommy is finally done!

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. BACKGROUND AND STATEMENT OF THE PROBLEM

Sleep and performance measures were collected during a previous study conducted in 2003 at Officer Indoctrination School (OIS) in Newport, Rhode Island. This thesis will analyze resulting ANAM Readiness Evaluation System (ARES), actigraphy, and sleep/activity log data. Analysis will include how ARES scores vary by subject, session, time of day, quality and quantity of sleep.

The actigraphy and sleep/activity log data have been interpreted, coded and imported into Fatigue Avoidance Scheduling Tool (FAST) to calculate subjects' predicted effectiveness. FAST is currently the preferred tool used to predict performance; it is based upon sleep debt from previous days, a sleep reservoir, and circadian oscillators. However, days of sleep and activity data must be collected before a meaningful assessment can be produced. In contrast, the ARES Commander Battery takes less than 10 minutes and can be administered on a digital personal assistant. ARES is a new software package that has not been validated, but is under consideration as a quick, inexpensive method of testing an individual's level of functioning in a military operational setting.

B. LITERATURE REVIEW

1. Sleep Deprivation and Performance Loss

Modern sleep research began in the mid-1950s with the discovery of two distinct states of sleep. Over the past 40 years, extensive research has been conducted on sleep, sleepiness, circadian rhythms, and sleep disorders, and how these factors affect waking alertness and performance

(Rosekind et al., 1996). Discussions of fatigue and subjective sleepiness and their relationship to alertness and performance occupy much of the literature. Although opinions differ, one subject matter expert gives the following definitions of fatigue, alertness and performance:

> . . . *Performance* comprises cognitive functions ranging in complexity from simple psychomotor reaction time, to logical reasoning, working memory and complex executive functions. By *alertness* is meant selective attention, vigilance, and attentional control. *Fatigue* refers to subjective reports of loss of desire or ability to continue performing. Additionally, subjective sleepiness is used [to describe] subjective reports of sleepiness or the desire to sleep (Van Dongen & Dinges, 2000, p. 2).

a. Military Research

Department of Defense funds research on the effects of sleep deprivation on human performance because military operations often provide limited sleep opportunities for many individuals. For example, the planned 96-hr SURGEOP on the USS NIMITZ required reduced sleep among personnel (Neri, Dinges, and Rosekind, 1997). Commanders need to know how long their crew can go without sleep before significant impairment. Captain David Neri, MSC, USN, Deputy Director of the Cognitive, Neural, and Biomolecular Science and Technology Division, Office of Naval Research writes about recent developments in modeling fatigue and performance:

> Stakes are high in the areas in which models are being used to inform, guide and confirm. These areas of current application include, but are not limited to: predicting individual and group performance; evaluating and guiding counter-measure use; schedule evaluation and design;

policy making (e.g., hours of service regulations); and accident assessment. For many in the operational community, biomathematical models of fatigue, sleepiness, and performance have become a significant issue. Military leaders, government policy makers, and commercial customers are looking for concrete answers to questions such as: how long can one work, fly or drive without rest or sleep; how much sleep is required for recovery; what is the minimum sleep necessary to sustain performance; when is a person most at risk for an error, incident, or accident; and what countermeasures can be taken at what time(s) to reduce these risks to an acceptable level? (Neri, 2004, p. A1)

b. Problems to Expect with Extended Sleep Deprivation

Sleep deprivation results in physiological and cognitive changes. Problems to expect include micro-sleeps, lapses in performance, reduced vigilance, poor communication, impaired decision making and short-term memory, and behavioral fixation. Additionally, sleep deprived individuals exhibit behavioral changes, such as slowed reaction times, increased errors and reduced performance on primary tasks. Degraded mood and reduced motivation have also been cited as deleterious effects due to sleep deprivation (Neri et al., 1997).

c. National Impact

The impact of sleep-related impairment is not limited to military operations. The 2001 Sleep in America Poll reported the prevalence of civilian sleep-related mishaps:

100,000 sleep-related car crashes per year;

1,500 fatalities

53% of adults report driving drowsy; 19%

dozed off at the wheel

27% report being sleepy at work at least 2 days/week

19% of adults report making errors at work; 2% injured

(National Sleep Foundation, 2001)

Several national disasters have been attributed to severe sleep deprivation. Two of these include the Exxon Valdez and Challenger incidents. On the night of March 24, 1989, the Exxon Valdez oil tanker ran aground, spilling millions of gallons of crude oil into the Prince William Sound. The cleanup cost was over $2 billion, leaving incalculable environmental damage. Additionally, Exxon Corporation was assessed $5 billion in punitive damages. While the media focused on the Captain’s alcohol consumption, the National Transportation Safety Board (NTSB) found that sleep deprivation was the direct cause of the accident (Dement & Vaughan, 1999). The following is an excerpt from Dement and Vaughan (1999):

> The report noted that on the March night when the Exxon Valdez steamed out of Valdez [, Alaska] there were ice floes across part of the shipping lane, forcing the ship to turn to avoid them. The captain determined that this maneuver could be done safely if the ship was steered back to the main channel when it was abeam of a well-known landmark, Busby Island. With this plan established, he turned over command to the third mate and left the bridge. Although news reports linked much of what happened next to the captain’s alcohol consumption, the captain was off the bridge well before the accident. The direct cause of America’s worst oil spill was the behavior of the third mate, who had slept only 6

> hours in the previous 48 and was severely sleep deprived.
>
> As the Exxon Valdez passed Busby Island, the third mate ordered the helm to starboard, but he didn’t notice that the autopilot was still on and the ship did not turn. Instead it plowed farther out of the channel. Twice lookouts warned the third mate about the position of lights marking the reef, but he didn’t change or check his previous orders. His brain was not interpreting the danger in what they said. Finally he noticed that he was far outside the channel, turned off the autopilot, and tried hard to get the great ship pointed back to safety—too late (p. 52).

Another national tragedy was the explosion of the space shuttle Challenger. The Rogers Commission investigation concluded that the decision to launch the rocket was an error given the inadequate data on O-ring function at low temperatures. However, according to Dement and Vaughan (1999), a less publicized fact is that the Human Factors Sub-committee cited severe sleep deprivation of the NASA managers as the cause of the error.

One may fault the employee(s) for not alerting their co-workers or supervisor to their impaired condition. However, research suggests that humans are not good at assessing their own impairment. Sagaspe (2003) led a study on fatigue, sleep restriction, and performance in automobile drivers. Simple reaction time, prospective self-assessment of performance, and instantaneous fatigue and sleep ratings were measured at two-hour intervals in both a sleep laboratory and on the open French highway. Under conditions of sleep restriction, some drivers took longer to brake in the natural environment than in the laboratory—an average of 23 meters in breaking distance at a speed of 75 miles per hour. A linear correlation between

self-assessment and reaction time was found in the laboratory condition but not in the road conditions. The researchers concluded that “The lack of correspondence between reaction time and prospective self-evaluation of performance suggests that self-monitoring in real conditions is poorly reliable” (Sagaspe, 2003, p. 277). Researchers at the Flight Management and Human Factors Division of NASA Ames Research Center would agree:

> A person is not a reliable judge of his or her own level of biological sleepiness. Careful studies using physiological measures of sleepiness have shown that people report a high level of alertness during the day and yet still exhibit significant physiological sleepiness. . . . Therefore, in attempting to judge how sleepy an individual is, the worst person to ask is that individual. It is better to rely on other signs and symptoms of fatigue that are related to performance decrements (Neri et al., 1997, p. 11).

2. Sleep Debt

According to Dement (2000), the average individual needs one hour of sleep for every two hours awake, which equates to eight hours per day. However, some individuals need more sleep and some need less, but each person has a specific daily sleep requirement. Supporting evidence comes from a recent sleep debt experiment conducted on 36 healthy subjects who spent 20 days inside a laboratory undergoing performance testing and restricted sleep (Van Dongen, Rogers, & Dinges, 2003). The study revealed that subjects’ estimated sleep need was 8.2 hours per day and the estimated standard deviation for interindividual differences in sleep need was 2.6 hours (Van Dongen, Rogers, & Dinges, 2003).

How people recover from lost sleep is still being studied. Thus far evidence suggests it must be paid back, possibly hour for hour (Dement & Vaughan, 1999). Mary Carskadon and William Dement use the term “sleep debt” to liken hours of required but unattained sleep to a monetary debt which must be paid back.

> Regardless of how rapidly it can be paid back, the important thing is that the size of the sleep debt and its dangerous effects are definitely directly related to the amount of lost sleep. My guess is that after a period of substantial sleep loss, we can pay back a little and feel a lot better, although the remaining sleep debt is still large. The danger of an unintended sleep episode is still there. Until proven otherwise, it is reasonable and certainly safer to assume that accumulated lost sleep must be paid back hour for hour (Dement & Vaughan, 1999, p. 60).

Sleep debt accumulates not only as a result of too few sleeping hours, but also from interrupted sleep. Sleep researchers have found that hundreds of nocturnal awakenings in a single night, despite normal cumulative amounts of total sleep, result in markedly increased daytime sleepiness (Dement & Vaughan, 1999).

Experiments on healthy adults, sleep restricted for six or more days, yielded

> statistically significant effects on daytime sleep latency [sleep onset], on daytime behavioral alertness as measured by psychomotor vigilance performance [PVT] lapses, on morning metabolic responses, on endocrine functions and on immune functions. Moreover, it appears that the sleep latency and behavioral alertness effects are directly related to the accumulation of sleep debt across days of sleep restriction (Van Dongen et al., 2003, p. 7).

Worth noting, a sleep-dose-dependent relationship between cumulative sleep debt and psychomotor vigilance tasks was revealed, but within the same study, waking electroencephalography (EEG) did not show progressive deterioration with additional sleep debt (Van Dongen et al., 2003). Apparently not all measures of waking function are good at identifying individuals’ sleep debt.

3. Sleep Regulation

Sleep debt can accumulate in small increments over days, such as during the work week, but, according to Dement and Vaughan (1999) it is difficult to pay back a sizeable debt over the weekend because of the biological clock’s alerting process. The biological clock regulates sleeping and waking to be in accordance with the daily rising and setting of the sun and seasonal light fluctuations. It also synchronizes biochemical events, such as chemical, hormonal, and nerve cell activities that influence daily fluctuations in feelings and actions (Dement & Vaughan, 1999). In an excerpt from *The Promise of Sleep*, Dement explains the competition between humans’ sleep drive and biological clock:

> The biological sleep drive that causes us to fall asleep and to remain asleep through the night is continuously active, even when we are awake. In fact, when we are awake the homeostatic sleep drive is steadily increasing. Opposing this sleep tendency is the alerting action of the biological clock. For humans and other diurnal animals, the clock-dependent alerting process is active in the daytime and inactive at night, with lowered activity in the early afternoon. The push and pull of these opposing processes allows us to stay up all day and sleep all night. In summary, the main reason we do not fall asleep as soon as we have been awake for a few hours is that the homeostatic sleep drive is held at bay

by the independent internal stimulation of the biological clock. The main reason that we can sleep through the night is that we have accumulated sufficient sleep debt during the day so that the unopposed homeostatic sleep process is free to operate all night long (p. 80).

The push and pull between the two internal regulators results in cycles of human wakefulness. Below is a graph depicting a simplified version of an individual's 24-hour alertness cycle. Other researchers have since labeled the two regulators: the homeostatic process and the circadian process.

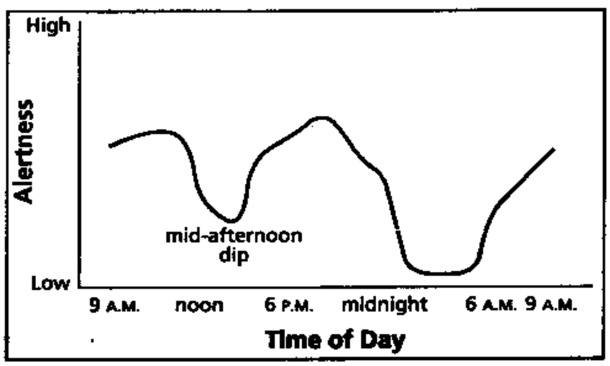


Figure 1. Homeostatic and Circadian Processes.
[From Mass, Wherry, Hogan, & Axelrod, 1998]

Variations of the two-process model of sleep regulation are used to predict the timing and duration of sleep. Van Dongen (2003) tested the model in a sleep debt experiment, described previously. The model predicts that chronic partial sleep deprivation will result in sleep-dose-related increases in homeostatic pressure. Within a few days, however, the average predicted waking homeostatic pressure stabilizes, suggesting adaptation to chronic sleep deprivation (Van Dongen et al., 2003).

Additionally, they examined whether the two-process model would predict neurobehavioral functioning. The

difference between predicted homeostatic pressure and observed PVT performance lapses were calculated relative to baseline for each individual. Analysis showed that the model did not predict neurobehavioral performance capability. The results also confirmed that sleep debt can lead to different responses depending on the measure of waking function (Van Dongen et al., 2003).

The circadian-homeostatic process model of sleep regulation appears to be missing a third unidentified process affecting waking behavioral alertness. Already identified are interindividual sleep need differences. Additionally, using waking EEG as a physiological marker of sleep homeostasis, Van Dongen (2003) found that naturally short sleepers tolerate a higher homeostatic pressure for sleep than long sleepers, suggesting a genetic basis for this variability in sleep need. Another source of natural variability, called vulnerability to sleep loss, is the differing magnitude of performance loss among individuals experiencing the same quantity of lost sleep. Using this additional knowledge, a linear mixed-effects model was applied to PVT performance deficits. When including inter-individual variability in ‘sleep need’ and ‘vulnerability to sleep loss’ in the model, 82.6% of the variance was explained by interindividual differences. In comparison, when the random effects were absent from the model, the explained variance dropped to 21.9%. “Thus, under conditions of chronic sleep restriction, sleep debt may be defined as the cumulative hours of sleep loss with respect to the subject-specific daily need for sleep” (Van Dongen et al., 2003, p. 11).

Another interindividual difference relates to the tendency to be a “lark” or an “owl”, that is, a morning or evening person. “Morning- and evening-type individuals differ endogenously in the circadian phase of their biological clock” (Kerkhof & Van Dongen, 1996, p. 153). Some people are consistently at their best in the morning, whereas others are more alert and perform better in the evening.

The three-process model of alertness is a recent expansion of the two-process model of sleep-wake regulation described earlier. Sleep inertia is the third process. Sleep inertia is the performance impairment and the feeling of disorientation experienced immediately after waking up. Studies have reported it to last from one minute to four hours with severity related to the duration of prior sleep. Sleep stage prior to awakening appears to be the most critical factor.

> Abrupt awakening during a slow wave sleep (SWS) episode produces more sleep inertia than awakening in stage 1 or 2, REM sleep being intermediate. Therefore, prior sleep deprivation usually enhances sleep inertia since it increases SWS. There is no direct evidence that sleep inertia exhibits a circadian rhythm. However, it seems that sleep inertia is more intense when awakening occurs near the trough of the core body temperature as compared to its circadian peak (Tassi & Muzet, 2000, p. 341).

4. Arousal and Alertness

According to Dement, the . . . “level of daytime alertness is probably the number-one determinant of how we will function mentally—learning, school performance, everything . . .” (Dement & Vaughan, 1999, p. 55).

> In the early days of sleep research, rather than talk about sleepiness or alertness itself, researchers measured the ability of sleep-deprived people to perform a task, such as stacking blocks in the right order or solving word puzzles. They called this measure 'performance failure' or 'fatigue.' The problem with this approach is that a person faced with a task can temporarily shake off fatigue. . . . Sleep-deprived test subjects presented with a task changed the conditions of the test by arousing themselves and masking the severity of their sleepiness, the very thing that researchers were trying to measure (Dement & Vaughan, 1999, p. 56).

Individuals often feel awake despite large sleep debts because sleepiness is counteracted by arousal. In addition to the biological clock, excitement or stress has alerting effects. While Dement notes that the effects of large sleep debt can be overcome in the short term by stimulating activities, recent studies suggest there is more to the matter. Research on heat loss and sleepiness (Matsumoto, Mishima, Satoh, Shimizu, & Hishikawa, 2002) found that among sleep deprived volunteers, physical exercise alleviated subjective sleepiness depending on the magnitude of the core body temperature elevation. However, performance still decreased, alerting him to the possibility . . . "that increased physical activity during extended wakefulness could increase the dissociation between subjective evaluation of sleepiness and actual brain function, resulting in increased risk of human error" (Matsumoto et al., 2002).

The U.S. Army Aeromedical Research Laboratory also examined the effectiveness of exercise for sustaining performance. The study consisted of two sessions. During the first session, participants engaged in ten minute bouts

of exercise throughout a 40-hour period of sleep deprivation. During the second session participants rested. Compared with the resting session, participants were more alert immediately following exercise, as evidenced by longer sleep latencies. However, “electroencephalogram data collected 50 minutes following exercise or rest showed that exercise facilitated increases in slow-wave activity, signs of decreased alertness. Cognitive deficits and slowed reaction times associated with sleep loss were equivalent in both conditions” (Le Due, Caldwell, & Ruyak, 2000, p. 249). Both studies concluded that exercise improves alertness, at least subjectively, but does not prevent performance decrements.

Other research indicates sustained performance under conditions of sleep deprivation is instable, perhaps explaining the differences in literature on arousal’s effect on alertness. Sleep deprivation does not eliminate the ability to perform neurobehavioral functions, but it does make it difficult to maintain stable performance for more than a few minutes. In a study investigating the variability in performance as a function of sleep deprivation, PVT reaction time means and standard deviations increased markedly among subjects and within each individual subject in the total sleep deprivation (TSD) condition relative to the 2-hour nap every 12 hours (NAP) condition (Doran, Van Dongen, & Dinges, 2001).

> Errors of omission [i.e., lapses] and errors of commission [i.e., responding when no stimulus was present] were highly intercorrelated across deprivation in the TSD condition, suggesting that performance instability is more likely to include compensatory effort than a lack of motivation. The marked increases in PVT performance

> variability as sleep loss continued supports the 'state instability' hypothesis, which posits that performance during sleep deprivation is increasingly variable due to the influence of sleep initiating mechanisms on the endogenous capacity to maintain attention and alertness, thereby creating an unstable state that fluctuates within seconds and that cannot be characterized as either fully awake or asleep (Doran et al., 2001, p. 253).

5. Sleep, Activity, Fatigue and Task Effectiveness Model (SAFTE™) and Fatigue Avoidance Scheduling Tool (FAST)

Principal investigator, Dr. Stephen Hursh at Science Applications International Corporation (SAIC) teamed up with talents from the Air Force Research Laboratory (ARFL), Walter Reed Army Institute of Research (WRAIR), and Federal Railroad Administration to develop software to manage fatigue and alertness for the operational components of the Services. Under an Air Force SBIR awarded to NTI, Inc., the software was developed and named Fatigue Avoidance Scheduling Tool (FAST). FAST is an actigraph-based application of the Sleep, Activity, Fatigue, and Task Effectiveness (SAFTE™) Model, developed by Hursh in 1996, but since modified. SAFTE™ is a three-process, quantitative model that was optimized to predict cognitive performance, rather than alertness (Eddy & Hursh, 2001). The following explanation of the Model comes from a paper circulated at the Fatigue and Performance Modeling Workshop held in Seattle, WA, June 2002, now published in *Aviation, Space and Environmental Medicine* (March 2004):

> The conceptual architecture of the SAFTE Model is shown in Figure [2]. The core of this model is schematized as a sleep reservoir, which represents sleep-dependent processes that govern the capacity to perform cognitive work. Under

fully rested, optimal conditions, a person has a finite, maximal capacity to perform, annotated as the reservoir capacity (Rc). While one is awake, the actual 'contents' of this reservoir are depleted, and while asleep, they are replenished. Replenishment (sleep accumulation) is determined by sleep intensity and sleep quality. Sleep intensity is in turn governed by both time-of-day (circadian process) and the current level of the reservoir (sleep debt). Sleep quality is modeled as its continuity, or conversely, fragmentation, in part determined by external, real-world demands, or requirements to perform. Performance effectiveness is the output of the modeled system. The level of effectiveness is simultaneously modulated by time-of-day (circadian) effects and the level of the sleep reservoir. Transient post-sleep decay of performance is modeled by the term inertia (Hursh, et al., 2004, p. A45).

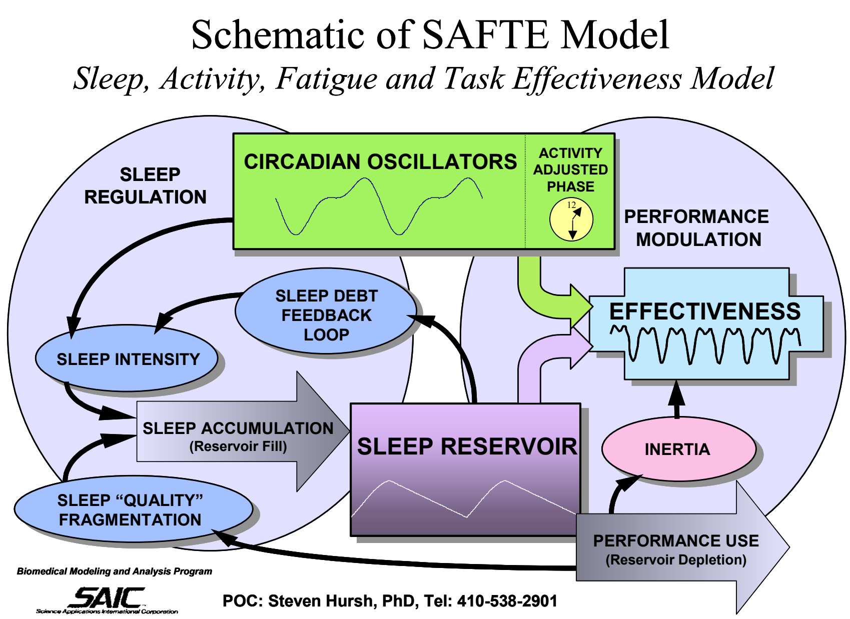


Figure 2. SAFTE™ Model. [From Eddy & Hursh, 2001]

In SAFTE™, cognitive performance capacity declines linearly during continuous wakefulness at a rate of about

1% per hour awake. “The rationale for both linearity and the value for the decay slope . . . is derived from a straight-line fit of cognitive throughput data obtained during 72 h of total sleep deprivation” (Hursh et al., 2004, p. A46). Additionally, the model estimates the circadian process as a two-frequency function. The circadian process is represented as the sum of two cosine waves, one with a period of 24 hours, the other with a period of 12 hours.

> The two oscillations are out of phase, producing an asymmetrical wave form: a gradual rise during the day with a plateau in the afternoon and a rapid decline at night that closely parallels published studies of body temperature. The circadian rhythm of performance is not a simple mirror image of variations in body temperature. The asymmetrical circadian rhythm combines with a gradually depleting reservoir process resulting in a bimodal variation in cognitive effectiveness that closely parallels published patterns of performance and alertness (Hursh et al., 2004, p. A47).

The developers of the SAFTE™ Model recognize its shortcomings:

> Two major limitations are that the model does not provide an estimate of group variance about the average performance prediction and it does not incorporate any individual difference parameters, such as age, morningness/ eveningness, or sleep requirement for full performance (Hursh et al., 2004, p. A51).

The importance of these limitations depends on how the model is applied. Using the model to predict a particular person’s fitness for duty is subject to higher predictive error than using the model to predict how a group will perform (Hursh et al., 2004). Others have found the importance of inter-individual differences to be more

important, explaining more than 50% of total variance in performance deficits resulting from up to 40 hours of sleep loss (Van Dongen, Maislin, & Dinges, 2004).

Another limitation of the SAFTE Model is that it does not account for the effects of pharmacological countermeasures, such as stimulants, used to extend performance or sedatives taken to enhance sleep. Stimulants can temporarily improve performance in sleep deprived individuals, but they can also interfere with sleep (Hursh et al., 2004).

Critics of the SAFTE model state that it requires validation in the field and modification in some areas. Although a validation study with the Department of Transportation Federal Railway Administration is planned, the model has not been validated outside the laboratory (Kronauer & Stone, 2004). Also, in comparison of mathematical model predictions to experimental data, the SAFTE model "in general did not predict performance well" (Van Dongen, 2004, p. A122). Commentary from the Fatigue and Performance Modeling Workshop concluded:

> [1]Although the 12-h circadian component was generally felt to be unnecessary, it was the linear function in performance decay that most of the audience found unacceptable. The concept of zero performance is not supported by experimental data (Kronauer & Stone, 2004, pp. A55-A56).

[1] In Response to Commentary on Fatigue Models for Applied Research in Warfighting, SAFTE developers "attempt to update and correct some of those impressions, based on the version of the model used at the Seattle conference, and respond to other concerns about the specific mathematical form of some of the model components" (Hursch & Balkin, 2004, p. A57).

As previously stated, SAFTE™ was applied in the development of FAST, a computerized tool to manage fatigue and performance. FAST was originally designed to help optimize the operational management of aviation ground and flight crews, although it is not limited to that application. FAST predicts performance effectiveness from sleep and work-schedule information. Corresponding Blood Alcohol Equivalencies are also given. Note that the majority of states consider driving with a blood alcohol level at or above .08 (grams per 10 deciliters) illegal. According to FAST, that blood alcohol level corresponds to a FAST performance effectiveness of 85%. Effectiveness at or above 90% is expected in individuals regularly receiving 8 hours of continuous sleep per 24 hour period. Effectiveness below 65% is expected to be critically impaired (Eddy & Hursh, 2001).

6. Automated Neuropsychological Assessment Metrics (ANAM) and ANAM Readiness Evaluation Tool (ARES)

Automated Neuropsychological Assessment Metrics 2001 (ANAM™ 2001) is a Windows-based system consisting of computerized tests and batteries designed for clinical and research applications. The tests were constructed to measure cognitive processing efficiency in a variety of psychological assessment contexts that include neuropsychology, fitness for duty, nuerotoxicology, pharmacology, and human factors research (Reeves, Winter, Kane, Elsmore, & Bleiberg, 2002). Subtests in ANAM™ are designed to “assess attention and concentration, working memory, mental flexibility, spatial processing, cognitive processing efficiency, memory recall, and arousal/fatigue level” (Reeves et al., 2002). Output includes accuracy, speed, and efficiency measures. Validation studies have

demonstrated that ANAM measures assess aspects of working memory, processing speed, and recall (Reeves et al., Draft 2002).

ARES (ANAM™ Readiness Evaluation System) consists of a subset of ANAM™ tests and was developed for use on handheld computers, such as Personal Digital Assistants (PDAs). The ARES Commander Battery is intended to provide operational commanders with an on-line assessment of a crewmember's ability to sustain levels of concentration, working memory, and mental efficiency. Although it was originally intended for commanders in command and control centers, it can be used in other military missions, such as sustained flight operations, to assess flight crew alertness and readiness (Elsmore & Reeves, 2002).

Data output includes the number of correct responses, mean and median response times, and throughput, a measure that represents both speed and accuracy in a single score. Throughput is computed as the average number of correct responses per minute during a testing session.

C. SCOPE, LIMITATIONS AND ASSUMPTIONS

Twenty newly-commissioned staff corps officers attending Officer Indoctrination School (OIS) volunteered for a study in 2003 conducted by Naval Postgraduate School (NPS) Information Technology graduate students developing standardized data collection and storage methods for Dr. Nita Miller of NPS. The study ran for five days, with each participant keeping a sleep/activity log, wearing an Actigraph wristwatch, and taking the ARES Commander Battery test on their personal digital assistant (PDA) three times per day. The rank of participants ranged from O-1 to O-3, ages 24-36, and consisted of twelve men and eight women,

all presumably healthy with no apparent sleep disorder. Participants experienced mild to moderate sleep deprivation during the normal course of their training.

II. METHOD

A. PARTICIPANTS

The participants included twenty volunteers, 12 males and eight females, ages 24 - 36. They were presumably healthy, with no apparent sleep disorders. Participants were recently commissioned staff corps officers with a minimum of 16 years of education and were of ranks O-1 through O-3.

B. APPARATUS AND INSTRUMENTS

Upon arrival, OIS distributed palm pilots on which the NPS researchers loaded Sleep and Activity Logs, and the ANAM Readiness Evaluation System (ARES). Three different ARES tests are available. The OIS study utilized the ARES Commander Battery, which measures Simple Reaction Time (a measure of basic psychomotor speed), Running Memory Continuous Performance Task (CPT) (a measure of working memory and executive functions), and administers the Stanford Sleepiness Scale (a subjective measure of alertness/fatigue). Additionally, participants wore actigraphs, a wristwatch-like device with an accelerometer that measures motion and is used to determine activity levels.[2]

C. DESIGN AND PROCEDURE

The study design is a prospective study, correlational in nature, with repeated measures of participants. Unlike a traditional analysis of variance (ANOVA), in which individuals are assigned randomly to different treatment

[2] For a thorough description of the methods employed, please refer to the NPS thesis written by O'Connor and Pattillo (December 2004). The study is described in Chapter VI. Naval Officer Indoctrination School Study in *Reengineering Human Performance and Fatigue Research through Use of Physiological Monitoring Devices, Web-Based and Mobile Device Data Collection Methods, and Integrated Data Storage Techniques*.

groups and then effects are assessed, in a repeated measures design individuals are subjected to more than one treatment (Girden, 1992). In this study, repeated measurements were obtained from the volunteers over five days. Actigraph data were collected, along with sleep and activity logs, and used for input into FAST.[3] Participants logged critical changes in their state, in particular, for example, when they went down for sleep, woke up, took the watch off, and when they went on and off watch standing duty. Additionally, participants were instructed to take the ARES Commander Battery three times a day for five days. ARES testing took approximately ten minutes per session.

[3] O'Connor and Patillo explain the transformation of raw actigraphy data into FAST files, the scoring algorithms employed, and subjective decisions they made regarding data cleaning.

III. ANALYTICAL STRATEGY

A. VARIABLES

1. Response Variable

FAST Predicted Performance Effectiveness score is the continuous response variable. Observations include FAST scores and ARES test results matched by time and date. Table 1 lists the range and quantiles of participants' FAST scores. Thirty scores are excluded from the analysis because those observations are missing one or more ARES score (see "NA's", Table 1). A histogram depicts the distribution of FAST scores (Figure 3). As expected, FAST data are negatively skewed with an average predicted effectiveness of 90.7%.

*** Summary Statistics for data in: CRM.and.SRT.SPLUS.data ***

	FAST
Minimum:	72.510
1st Quantile:	87.210
Mean:	90.738
Median:	91.650
3rd Quantile:	94.990
Maximum:	101.530
Total N:	415.000
NA's:	30.000
Standard Deviation:	5.932

Table 1. Descriptive Statistics for FAST Performance Effectiveness.

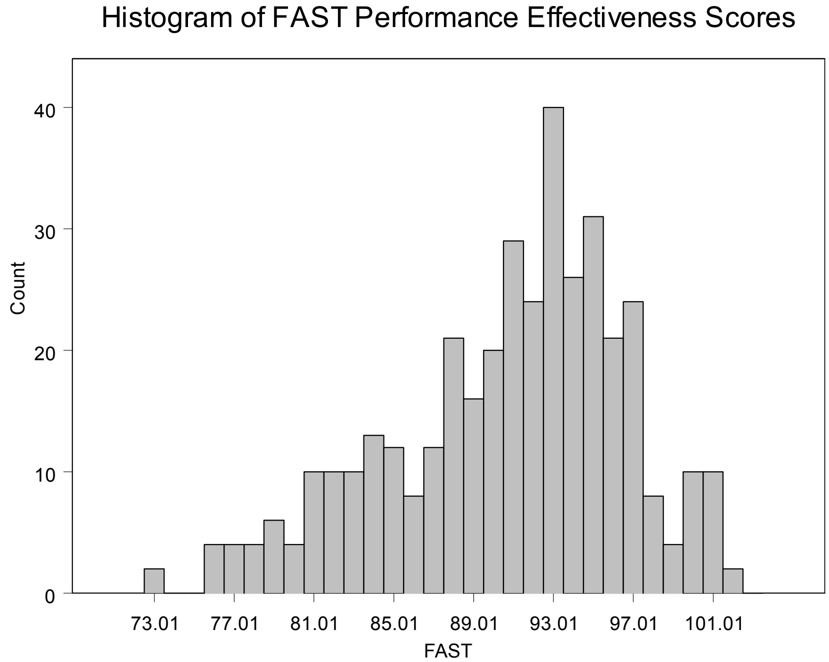


Figure 3. Histogram of Observed Fast Scores.

2. Predictor Variables

a. Time Blocks

FAST incorporates a circadian process within the SAFTE™ Model (see Figure 2). The Model's circadian oscillator is shown in Figure 4. Major peaks in performance and alertness are seen at about 1000 and 2000. Minimums are in the early afternoon, at about 1400, and in the early morning, around 0400. (Hursh, 2001)

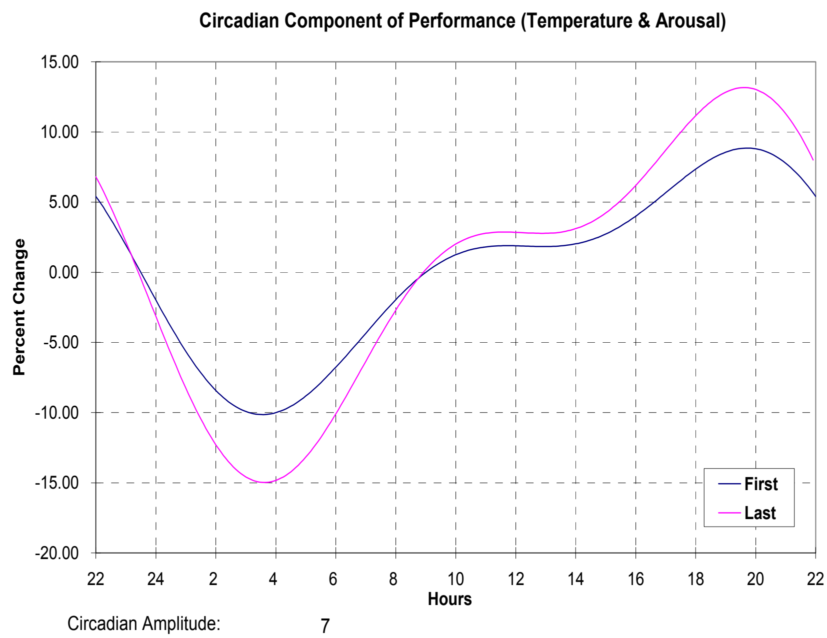


Figure 4. Circadian Oscillator in FAST. The Curve Marked First and Last Are for the First and Third Days, Respectively, of 72 Hours of Sleep Deprivation. [From Hursh, 2001]

Time blocks were created to reflect FAST's circadian oscillator and the OIS sleep plan. During the OIS study, unless assigned to the night watch, participants were allowed to sleep from 2200 to 0600. Various partitioning of the 24-hour day were explored in MS Excel. The following five partitions appeared to be significant, so the FAST and ARES scores were grouped according to these time blocks (Table 2). As expected, Table 3 shows that participants rarely took the ARES Commander Battery between midnight and 0437 (i.e., Time Block 1).

Time Block	From - To
1	00:00 - 04:47
2	04:48 - 09:35
3	09:36 - 14:23
4	14:24 - 19:11
5	19:12 - 23:59

Table 2. The 24-hour Day Partitioned into Five Equal Time Blocks, Each Four Hours and 48 Minutes Long, Starting at Midnight.

*** Summary Statistics for data in: CRM.and.SRT.SPLUS.data ***

Time.Block	Frequency
1:	6
2:	125
3:	118
4:	61
5:	105

Table 3. Number of Observations for Each Time Block.

b. Subject and Session

Although the OIS study had 20 participants, only two people completed all 15 scheduled ARES testing sessions (Figure 5). No test scores were collected for participant 6 and participant 15 tested only once. The average number of ARES sessions across participants is 6.43, and the standard deviation is 3.66. Subject is treated as a factor and Session is an integer.

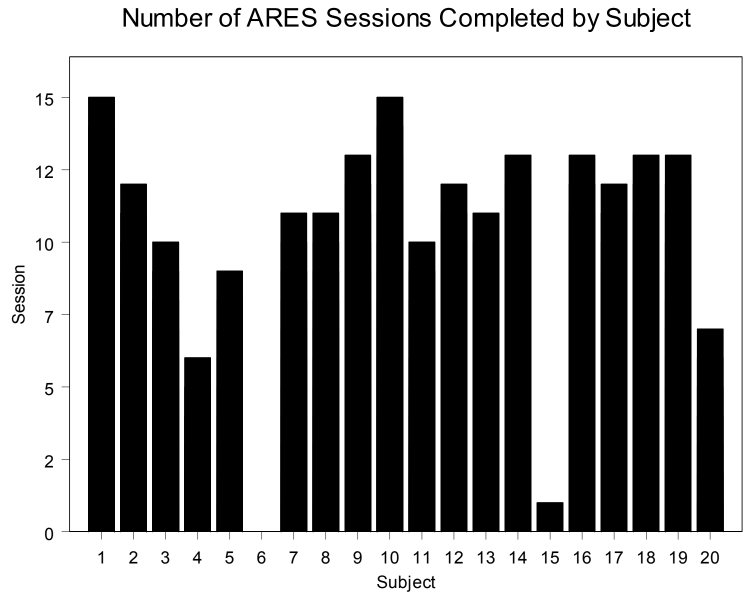


Figure 5. The Number of ARES Testing Sessions Recorded for Each Participant.

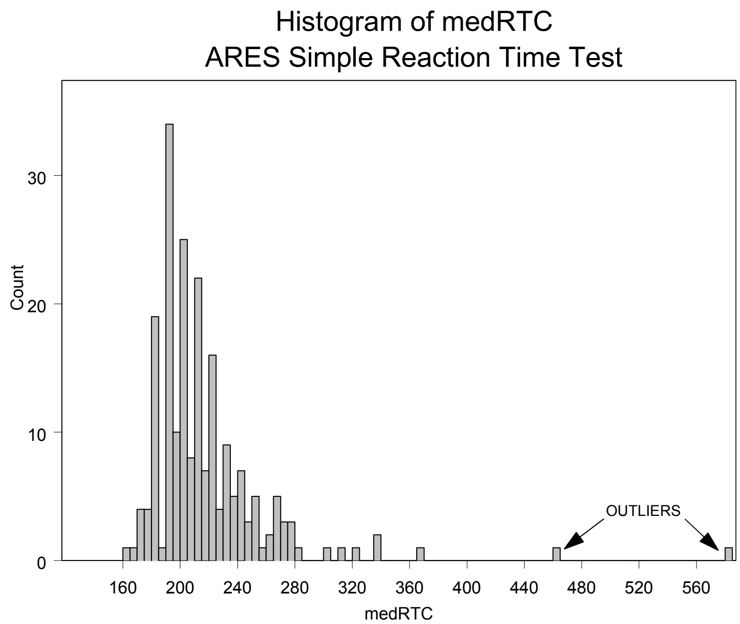
c. Simple Reaction Time

The median reaction time for correct responses (medRTC) and the standard deviation of reaction time for correct responses, 1st half of the testing session (sdRTC1), are continuous numeric variables. Observations for both variables are positively skewed (Figure 6). It is apparent that the two maximum values, 580 milliseconds for medRTC and 961 milliseconds for sdRTC1, are outliers; the majority of data fall close to the median (Table 4).

*** Summary Statistics for data in: CRM.and.SRT.SPLUS.data ***

	medRTC	sdRTC1
Minimum:	160.000	7.000
1st Quantile:	190.000	25.000
Mean:	215.553	56.947
Median:	205.000	40.000
3rd Quantile:	226.250	67.500
Maximum:	580.000	961.000
Total N:	208.000	208.000
NA's:	0.000	0.000
Standard Deviation:	44.421	80.280

Table 4. Range and Quantiles of the Median (medRTC) and Standard Deviation (sdRTC1) of Reaction Time for the ARES Simple Reaction Time Test.



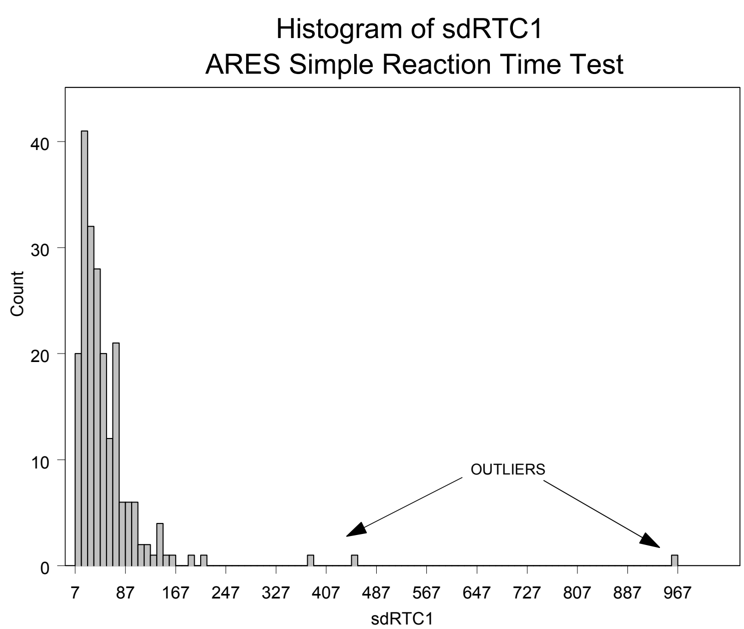


Figure 6. Distribution of the Median (medRTC) and Standard Deviation (sdRTC1) of Reaction Time Observations for the ARES Simple Reaction Time Test.

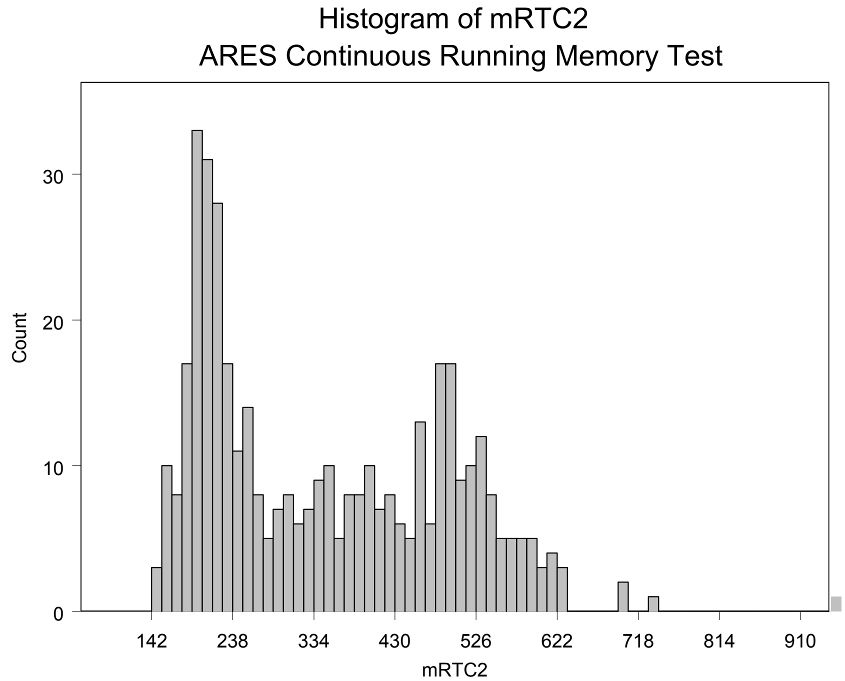
d. Continuous Running Memory

A continuous numeric variable, sdRTC2 is the standard deviation, in milliseconds, of the reaction time for correct responses during the second-half of the testing sessions. Also a numeric variable, mRTC2 is the mean reaction time of correct responses during the second-half of each session; it is the average response latency in milliseconds. Histograms illustrate the shape of the distributions of sdRTC2 and mRTC2 (Figure 7). SdRTC2 is negatively skewed and ranges from 39 to 190, with a mean of 115.6 and standard deviation of 34.3 (Table 5). MRTC2 is positively skewed and bimodal; observations range from 297 to 736, the mean is 464.5 and the standard deviation is 88.4 (Table 5).

*** Summary Statistics for data in: CRM.and.SRT.SPLUS.data ***

	sdRTC2	mRTC2
Minimum:	39.000	297.000
1st Quantile:	90.000	394.000
Mean:	115.551	464.473
Median:	119.000	472.000
3rd Quantile:	140.000	526.000
Maximum:	190.000	736.000
Total N:	207.000	207.000
NA's :	0.000	0.000
Standard Deviation:	34.343	88.391

Table 5. Descriptive Statistics for the Standard Deviation (sdRTC2) and Mean (mRTC2) of Reaction Time during the 2^{nd} half of the ARES Continuous Running Memory Test.



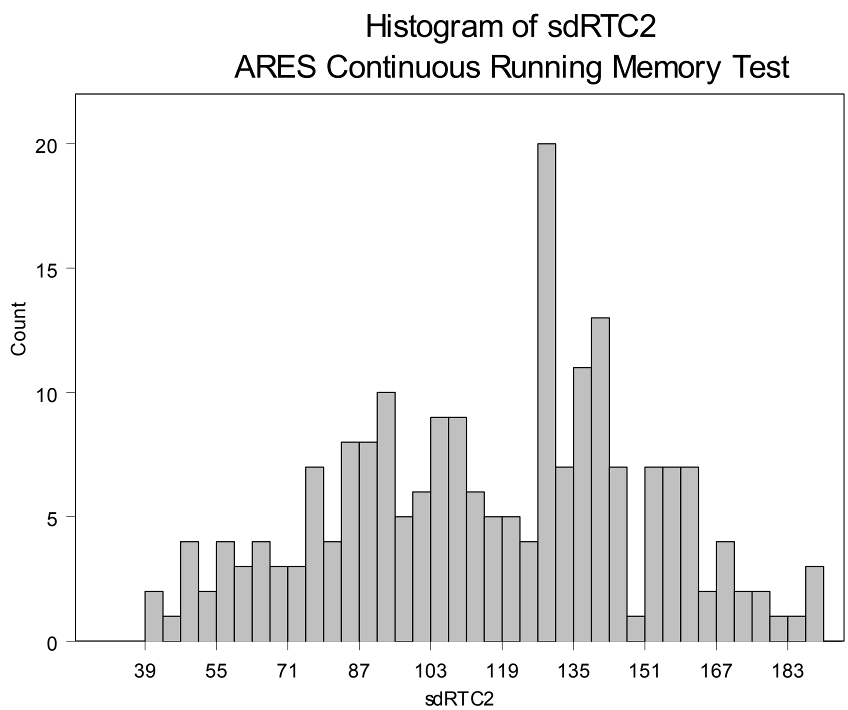


Figure 7. Distribution of the Mean (mRTC2) and Standard Deviation (sdRTC2) of Reaction Time Observations for the ARES Continuous Running Memory Test.

B. DESCRIPTIVE STATISTICS

The range and variability of reaction time for correct responses differ among OIS participants. As seen in Figure 8a, for some participants, the range of variability in reaction time is double that of co-participants (e.g., the sdRTC1 for Subject 17 is more than double that of Subject 18). MedRTC appears to be Subject-specific; each participant has his own distribution of reaction times, not necessarily overlapping other participants' observations. For example, Subjects 13 and 16 have no scores in common with Subjects 17 and 20 (Figure 8b).

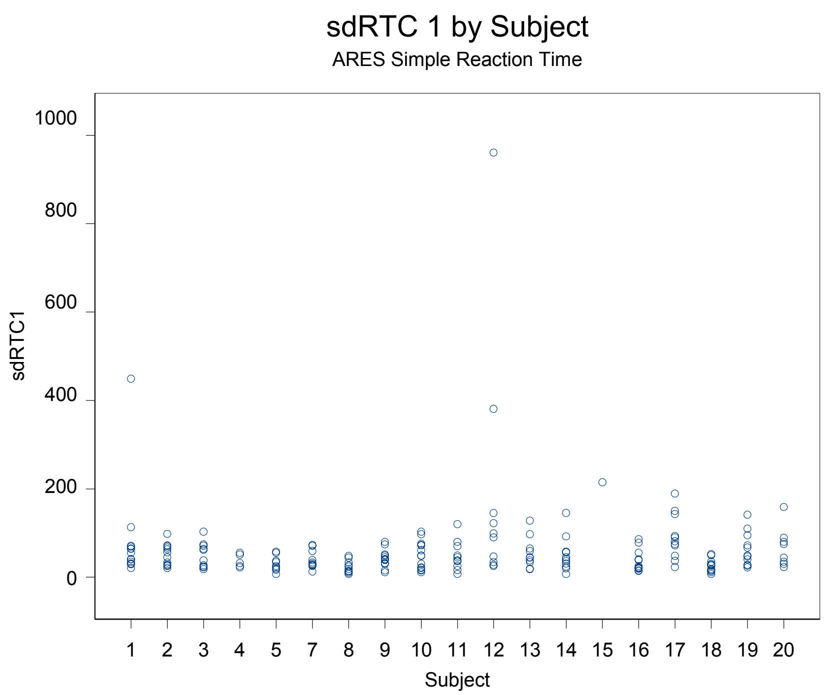


Figure 8a

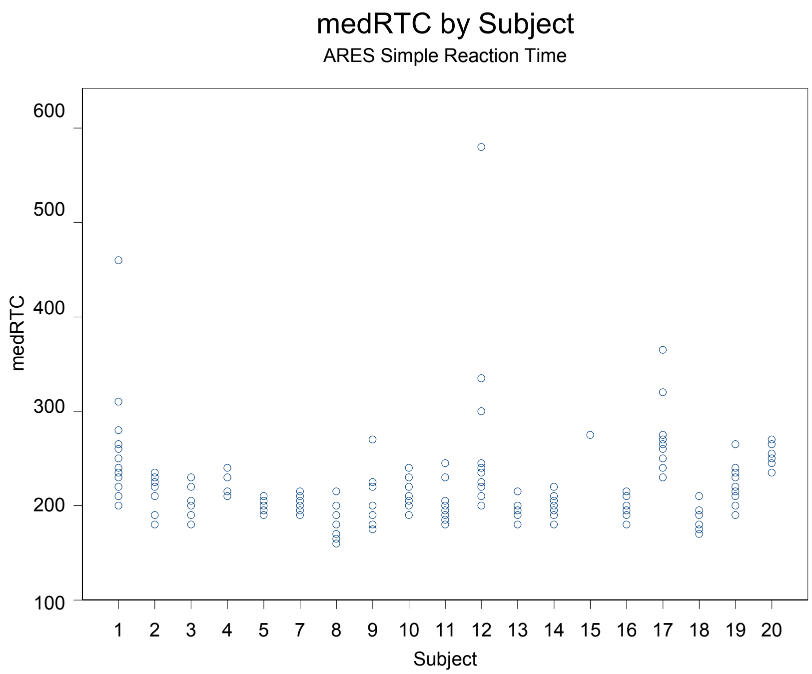
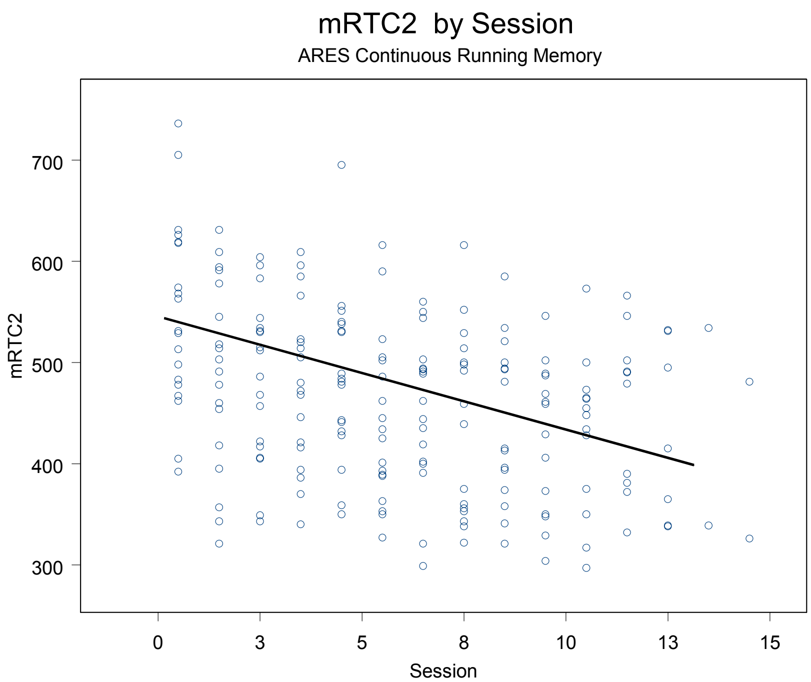


Figure 8b

Figure 8. Standard Deviation (sdRTC1) and Median (medRTC) of Reaction Time for Correct Responses by Subject

The ARES Continuous Running Memory predictor variables, standard deviation in reaction time (sdRTC2) and mean reaction time (mRTC2) for correct responses, are plotted against Session (Figure 9). MRTC2 has an obvious downward trend as the Session number increases; improvement in sdRTC2 is questionable. SdRTC2 seems to improve up through Session 7, after which the pattern is not apparent (Figure 9). Improvements across Session are suggestive of a practice-effect.



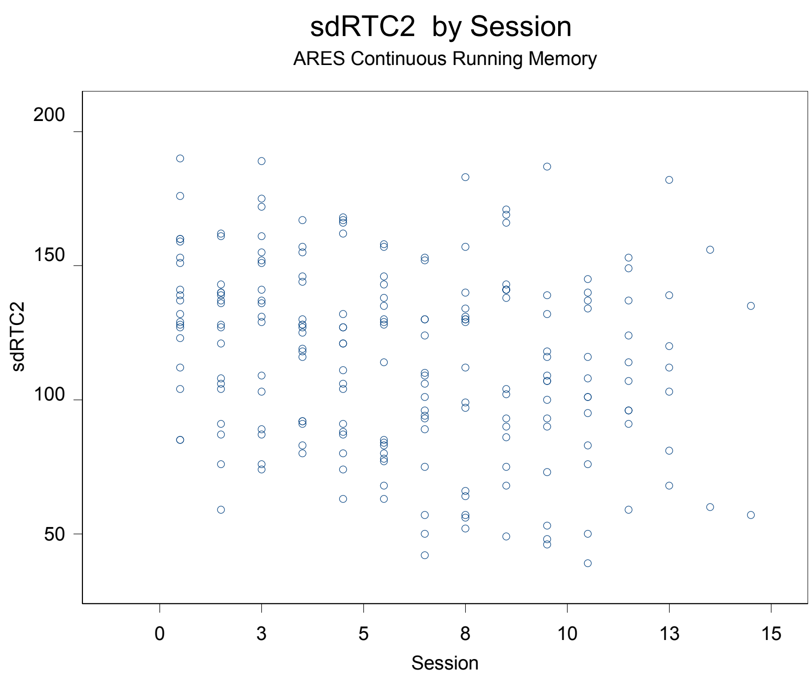


Figure 9. Mean (mRTC2) and Standard Deviation (sdRTC2) in Reaction Time for Correct Responses across Sessions.

C. REGRESSION MODEL AND ANALYSIS

A linear mixed-effect regression model was developed using S-PLUS 6.1, a statistical software package. (S-PLUS 6.1 for Windows Supplement, 2002) Mixed-effect models are appropriate for repeated measures data because they incorporate both fixed and random effects. Fixed effects are parameters associated with an entire population, or with repeatable levels of experimental factors. Random effects are associated with experimental units drawn at random from a population. The predictor variables are modeled as fixed effects, and their parameters are estimated by restricted maximum likelihood (REML). The Fixed-Effect part of the linear mixed-effect model assumes that the response, FAST scores, is obtained by taking a linear combination of the predictors. The within-group errors have a Gaussian (normal) distribution and are allowed to be correlated and/or have unequal variances (S-PLUS 2000 Professional Edition for Windows, Release 3, LME Help).

Two linear mixed-effect regression models are developed, one using the ARES Simple Reaction Time test data, the other using Continuous Running Memory test data (Figure 10). For ARES Simple Reaction Time data, the random effect is modeled by a random intercept and grouped by Subject. The random effect of ARES Continuous Running Memory is also modeled by a random intercept, but is grouped by Session. Time Block is a fixed effect common to both models. Additional fixed effect predictors for the Simple Reaction Time model are medRTC and sdRTC1. For the Continuous Running Memory model, sdRTC2 and mRTC2 are additional fixed effects.

a) Random effects: ~ 1 | Subject
Fixed: FAST ~ Time.Block + sdRTC1 + medRTC

b) Random effects: ~ 1 | Session
Fixed: FAST ~ Time.Block + sdRTC2 + mRTC2

Figure 10. Linear Mixed-Effect Model Formula for a) Simple Reaction Time, and b) Continuous Running Memory

THIS PAGE INTENTIONALLY LEFT BLANK

IV. RESULTS

A. ARES SIMPLE REACTION TIME LINEAR MIXED-EFFECTS MODEL

The linear mixed-effects model, using ARES Simple Reaction Time data, is FAST’ ~ 93.140 +6.148 * (Time.Block1) + 1.344 * (Time.Block2) + 1.374 * (Time.Block3) -1.117 * (Time.Block4) + 0.010 * (sdRTC1) - 0.020 * (medRTC). This is a regression prediction equation; it describes the prediction of FAST scores based on the predictor variables used in the regression analysis (i.e., the right side of the equation). The intercept and coefficients for each variable come from the statistical report in Figure 12 (see numbers under Value).

The intercept is 93.140. If values are unavailable for the predictor variables (i.e., they are set to zero in the equation), the model predicts a FAST performance effectiveness of 93.14%. Time.Block is a binary variable; its value can be zero or one. Valid values for SdRTC1 and medRTC are continuous numbers that fall within the range of data used to generate the model (i.e., 7 to 961 milliseconds for sdRTC1 and 160 – 580 milliseconds for medRTC). For example, if an individual takes the ARES Simple Reaction Time test at 1015 and his medRTC is 205 milliseconds and his sdRTC1 is 40 milliseconds, using the regression prediction equation, his predicted FAST score equals 90.814, or 90.81% (Figure 11).

$$FAST' \sim 93.140 + 6.148 * (Time.Block1) + 1.344 * (Time.Block2) + 1.374 * (Time.Block3) - 1.117 * (Time.Block4) + 0.010 * (sdRTC1) - 0.020 * (medRTC)$$

$$\text{Predicted FAST} = 93.140 + 6.148*(0) + 1.344*(0) + 1.374*(1) - 1.117*(0) + 0.010*(40) - 0.020*(205)$$

$$= 90.814 = 90.81\%$$

Figure 11. Computing a Predicted FAST Performance Effectiveness Score using the ARES Simple Reaction Time Linear Mixed-Effects Prediction Equation.

According to the statistical report (Figure 12), there is a high probability that there is a relationship between FAST performance effectiveness and the predictor variables (i.e., Time.Block, sdRTC1, and medRTC). The results are statistically significant, as evidenced by p-values less than .05. The .05 *p*-value is sufficiently stringent to safeguard against accepting too many insignificant results as significant, while not being overly difficult to attain (Newton & Rudestam, 1999).

```
                          *** Linear Mixed Effects Model ***
Random effects:
 Formula:  ~ 1 | Subject
                    (Intercept)  Residual
Standard Deviation:    3.069       3.427

Fixed effects: FAST ~ Time.Block + sdRTC1 + medRTC
```

	Value	Standard Error	Degrees of Freedom	t-value	p-value
(Intercept)	93.139	2.080	170	44.777	0.000
Time.Block1	6.148	1.047	170	5.873	0.000
Time.Block2	1.344	0.384	170	3.502	0.001
Time.Block3	1.373	0.249	170	5.509	0.000
Time.Block4	-1.117	0.154	170	-7.240	0.000
sdRTC1	0.010	0.004	170	2.315	0.022
medRTC	-0.020	0.010	170	-2.137	0.034

Standardized Within-Group Residuals:

Minimum	Quantile 1	Median	Quantile 3	Maximum
-2.283	-0.586	0.090	0.516	2.520

Number of Observations: 193   Number of Groups: 17

Figure 12. SPLUS 6.1 Report for ARES Simple Reaction Time Linear Mixed-Effects Model

Diagnostic plots displayed in Figure 13 indicate that modeling assumptions are met. A residual, or prediction error, is the difference between the actual and predicted FAST score. Prediction error is expected across the range of FAST scores, but variance must be constant (homoscedastic). As shown in Figure 13a, the ARES Simple Reaction Time linear mixed-effects model has homoscedastic residuals; they are scattered randomly. In contrast, heteroscedasticity is indicated when the residuals spread or fan out from left to right or right to left.

An additional assumption of linear regression is that within-group errors have a Gaussian (normal) distribution (i.e., a bell shaped curve that is symmetrical and unimodal). A normal probability plot or Quantile-Quantile (Q-Q) plot is used to evaluate whether or not the data meet this assumption. Figure 13b is a Q-Q plot for the ARES Simple Reaction Time model. The horizontal axis shows the location of the points as observed in the distribution. The vertical axis shows the location of the points as expected if the distribution is normal. A diagonal straight line, as seen in Figure 13b, indicates that the observed and expected distributions are the same (i.e., the distribution is normal), as required.

A final assumption of linear regression is the absence of correlation between error terms (i.e., how strongly they are related). This assumption is tested using an autocorrelation plot (Figure 13c), which displays the correlation of errors (i.e., residuals) across cases. The length of the vertical bars represents the magnitude of the correlation, with the value of +/- 1.0 indicating perfect correlation. However, the first position (i.e., Lag 0) is

always 1.0. Figure 13c shows that autocorrelation is acceptable for this model.

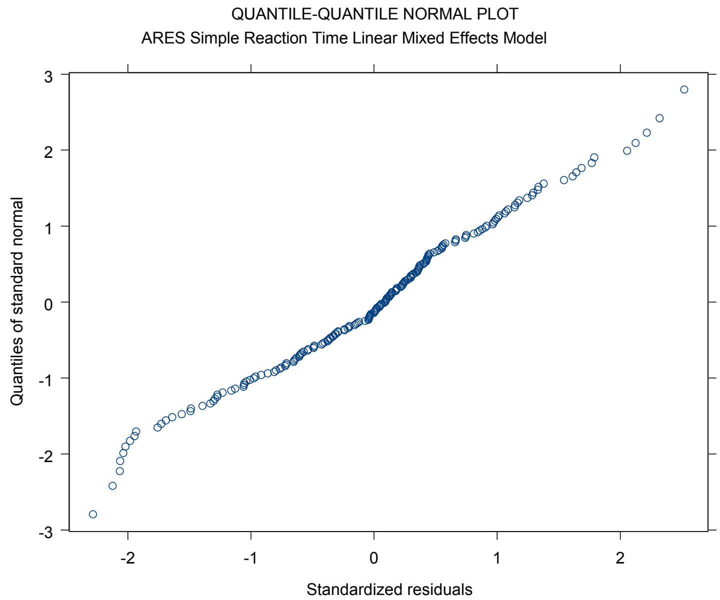


Figure 13a

Figure 13b

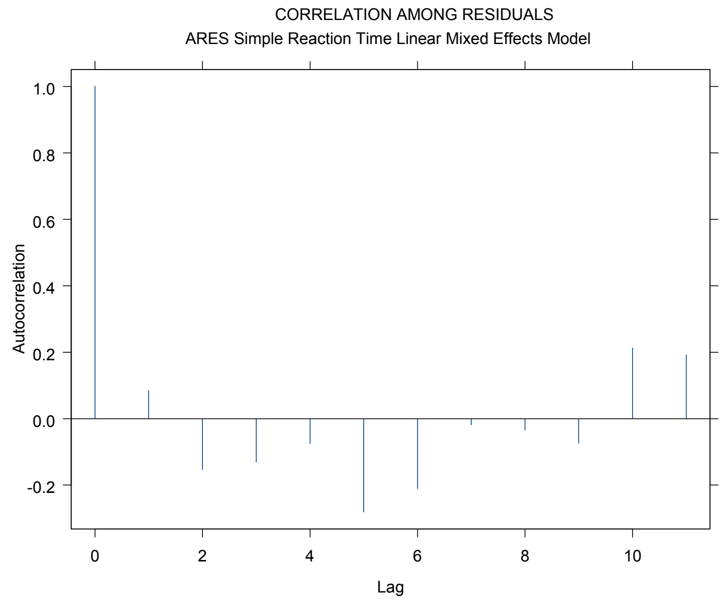


Figure 13c

Figure 13. ARES Simple Reaction Time Linear Mixed-Effects Model Diagnostic Plots: a) QQ-norm, b) Residuals vs. Fitted Values, c) Autocorrelation of Residuals

B. ARES CONTINUOUS RUNNING MEMORY LINER MIXED-EFFECTS MODEL

The linear mixed-effects model using ARES Continuous Running Memory data is FAST′ ~ 87.976 + 5.930 * (Time.Block1) + 1.180 * (Time.Block2) + 1.4884 * (Time.Block3) -0.983 * (Time.Block4) + 0.052 * (sdRTC2) - 0.010 * (mRTC2). The intercept and coefficients come from SPLUS 6.1 output (see Value, Figure 15) As with the previous regression prediction equation (i.e., for Simple Reaction Time), Time.Block variables can be either zero or one, with a one indicating the new observation falls within that time block. Also, valid input for sdRTC2 can be any continuous number between 39 and 190 milliseconds. For mRTC2, values must be between 297 and 736 milliseconds. The intercept is 87.976. If inputs are unavailable for the

predictor variables, the predicted FAST performance effectiveness is 87.98%. As an example, a new observation occurs at 1550, consisting of an ARES Continuous Running Memory mRTC2 of 472 milliseconds, and an sdRTC2 of 119 milliseconds, the predicted FAST score is 88.461, or 88.46% (Figure 14).

$$\text{FAST}' \sim 87.976 + 5.930 * (\text{Time.Block1}) + 1.180 * (\text{Time.Block2}) + 1.488 * (\text{Time.Block3}) - 0.983 * (\text{Time.Block4}) + 0.052 * (\text{sdRTC2}) - 0.010 * (\text{mRTC2})$$

$$= 87.976 + 5.930*(0) + 1.180*(0) + 1.488*(0) - 0.983*(1) + 0.052*(119) - 0.010*(472)$$

$$= 88.461 = 88.46\%$$

Figure 14. Computing a Predicted FAST Performance Effectiveness Score using the ARES Continuous Running Memory Linear Mixed-Effects Prediction Equation.

Additionally, the probability of a relationship between FAST performance effectiveness and the model’s predictor variables (i.e., Time.Block, mRTC2, and sdRTC2) is high. All Time Blocks and sdRTC2 are significant to the alpha < .05 level (Figure 15). The mRTC2 p-value is .06, but is retained in the model to encourage further exploration of the variable’s relationship with FAST.

```
                    *** Linear Mixed Effects Model ***

Random effects:
 Formula:  ~ 1 | Session
                    (Intercept)     Residual
Standard Deviation:    0.002         4.460

Fixed effects: FAST ~ Time.Block + sdRTC2 + mRTC2
```

	Value	Standard Error	Degrees of Freedom	t-value	p-value
(Intercept)	87.976	1.842	171	47.755	0.000
Time.Block1	5.930	1.322	171	4.485	0.000
Time.Block2	1.179	0.487	171	2.420	0.017
Time.Block3	1.488	0.307	171	4.844	0.000
Time.Block4	-0.983	0.191	171	-5.136	0.000
sdRTC2	0.052	0.014	171	3.717	0.000
mRTC2	-0.010	0.005	171	-1.868	0.063

Standardized Within-Group Residuals:

Minimum	Quantile 1	Median	Quantile 3	Maximum
-2.199	-0.676	0.107	0.629	3.058

Number of Observations: 192
Number of Groups: 15

Figure 15. SPLUS 6.1 Report for ARES Continuous Running Memory Linear Mixed-Effects Model

Diagnostic plots of the model’s residuals indicate that modeling assumptions are met. Residuals are homoscedastic (Figure 17a), within-group errors have a Gaussian (normal) distribution (Figure 16b), and there is no strong correlation among residuals (Figure 16c).

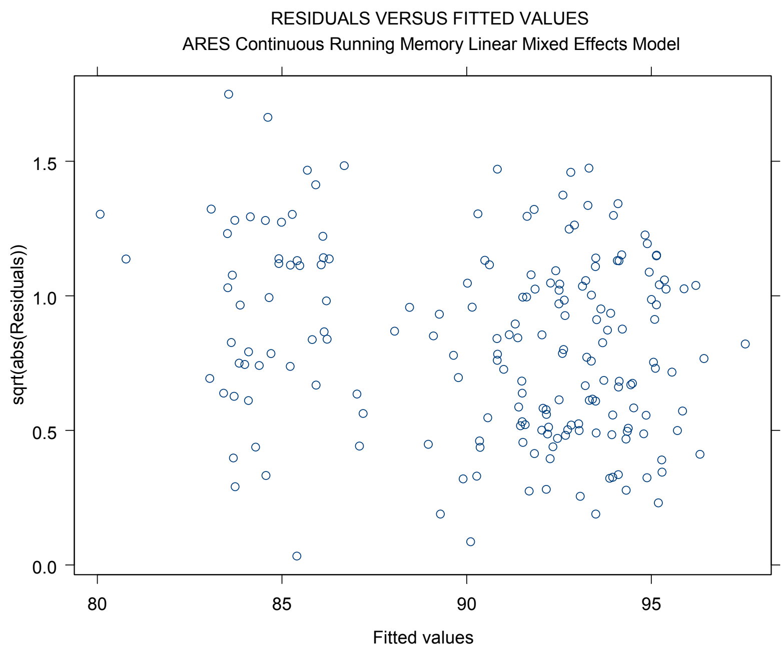


Figure 16a

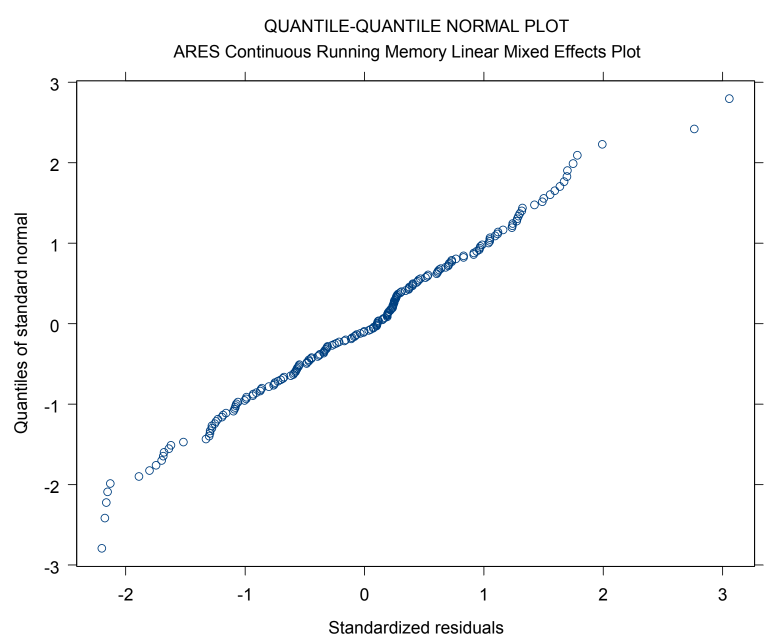


Figure 16b

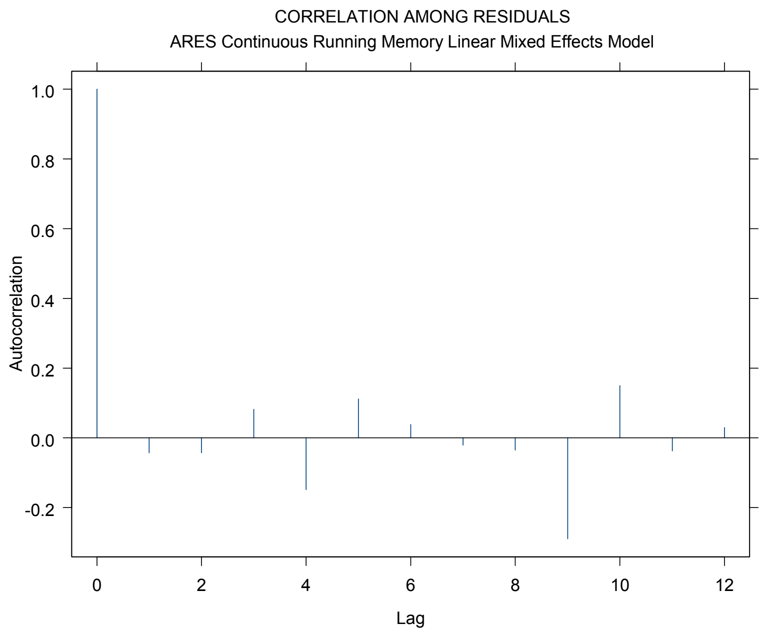


Figure 16c

Figure 16. ARES Continuous Running Memory Linear Mixed-Effects Model Diagnostic Plots: a) QQ-norm, b) Residuals vs. Fitted Values, c) Autocorrelation of Residuals

THIS PAGE INTENTIONALLY LEFT BLANK

V. DISCUSSION

Modeling fatigue, sleepiness, and performance is of significant interest to the military operational community. Because a person is not a reliable judge of his or her own level of biological sleepiness, commanders require an objective means to assess their crewmembers' ability to perform. One such method is FAST, the software application based upon SAFTE™. SAFTE™ is a biomathematical model designed to predict individual and group performance under conditions of sleep deprivation. Also, psychomotor vigilance tests, such as the ARES Commander Battery, provide instant feedback on an individual's ability to sustain levels of concentration, working memory, and mental efficiency.

FAST is currently the preferred tool used to predict performance. However, days of sleep and activity data must be collected before a meaningful assessment can be produced. In contrast, the ARES Commander Battery takes less than 10 minutes and can be administered on a digital personal assistant. ARES is a new software package that has not been validated, but is under consideration as a quick, inexpensive method of testing an individual's level of functioning in a military operational setting.

Analysis of Officer Indoctrination School data was aimed at identifying how ARES Simple Reaction Time and Continuous Running Memory test scores vary by subject, session, and time of day. Additionally, the relationship between ARES data and FAST performance effectiveness scores were explored. Time of day was partitioned into five time blocks that capture the changing direction of the human

alertness curve (see Figure 1). Linear mixed-effects models were built using search strategies, that is, all possible combinations of ARES variables were explored as predictors of FAST scores (i.e., the response variable). ARES variables analyzed include the mean, median, and standard deviation of reaction times for correct and incorrect responses; throughput, a measure of speed and accuracy; and, inter-trial responses, key presses between stimuli when the screen is blank. These measures were available for the entire session, the first half, and the second half of each trial.

Two linear mixed-effects models were developed; one using ARES Simple Reaction Time data, the second using ARES Continuous Running Memory data. Time Block was included as a fixed effect in both models. The standard deviation (sdRTC1) and median (medRTC) reaction time for correct responses are additional fixed effects in the ARES Simple Reaction time model (Figure 10). For the ARES Continuous Running Memory model, the standard deviation (sdRTC2) and mean (mRTC2) reaction time for correct responses are fixed effect predictor variables (Figure 10).

Mixed-effects modeling is preferred in research on human neurobehavioral functions because it allows for isolation of variability due to both inter- and intra-individual differences (Van Dongen et al., 2004). The ARES Simple Reaction Time linear mixed-effects model requires Subject in the random effects formula. Without Subject, the fixed effects predictors, with the exception of Time Block, were statistically insignificant. Additionally, the residuals were heteroscedastic and non-normal. Clearly, Subject must be modeled as a random effect.

For the ARES Continuous Running Memory linear mixed-effects model, Session was the key random effect. Subject was explored, but did not lead to a good model. It is important to note that these models are almost certainly over-fit to the OIS data. Numerous variations and combinations of predictor variables were explored. The final models include the only statistically significant combination of variables found to adhere to linear regression modeling assumptions. Because variable selection based on searching exploits chance patterns in the Officer Indoctrination School sample, conclusions should not be applied to other samples or the population. Additional studies need to be conducted to further explore these findings.

Additional insights came from in-depth exploration of variables. Unexpectedly, the three variables for throughput (i.e., throughput, throughput1, and throughput2) did not account for variance in FAST performance effectiveness. Also, many ARES scores, including those used in the models, improve with additional sessions, suggesting a potential bias posed by training. There is an indication that performance improves with continued trials in this study, a phenomenon commonly observed in human research.

An advantage of a repeated measures strategy is that it requires fewer individuals and the group serves as its own control. However, disadvantages include attrition of subjects. While this study started with 20 volunteers, only two participants completed all fifteen testing sessions. Also, practice, carry-over and fatigue can bias the results.[4] Evidence of a practice-effect is seen in the

downward, improving trend of ARES testing measures (e.g., mRTC2) as the number of testing sessions increase.

Overall, this study identified ARES variables that show promise as instantaneous indicators of human performance decrement under conditions of mild sleep deprivation (i.e., an average of six hours per night). Equally important, although it was initially expected for throughput to be the primary indicator of an individual's biological sleepiness, throughput did not account for variance in FAST performance effectiveness. Additionally, inter-individual differences accounted for much of the variability in ARES Simple Reaction Time scores, but Session explained variability in ARES Continuous Running Memory scores.

It is recommended that future studies include numerous practice sessions on the ARES Commander Battery to overcome the improving trend found across sessions. Additionally, in this study, baseline FAST performance effectiveness values were set to individuals' average FAST score during the five-day study. The three days prior to the study were conditioned, on an individual basis, to the average sleep time per night of the study. For example, if a participant averaged 362 minutes per night, this average was used to condition FAST for the three days prior to data collection. To ensure accurate baseline FAST performance effectiveness values, it is recommended that adequate actigraphy and sleep log data be collected prior to beginning the study data collection period.

[4] Girden (1992) discusses biases and methods to correct for bias, however most limitations of a repeated measures design appear to be an issue when multiple levels (i.e., more than one treatment) are employed. This OIS study uses only one level.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Belenky, G. (n.d.). “Sleep, Sleep Deprivation, and Human Performance in Continuous Operations.” Retrieved August 2003 from http://www.usafa.af.mil/jscope/JSCOPE97/ Belenky 97/Belenky97.htm.

Dement, W. C., & Vaughan, C., *The Promise of Sleep*, 52-57, 60-61, 76, 80, New York, Dell Publishing, 1999.

Doran, S. M., Van Dongen, H. P., & Dinges, D. F. (2001). “Sustained Attention Performance during Sleep Deprivation: Evidence of State Instability,” *Archives Italiennes de Biologie*, 139(3), 253-267.

Eddy, D. R., & Hursh, S. R. (2001). *Fatigue Avoidance Scheduling Tool (FAST).* AFRL-HE-BR-TR-2001-0140, SBIR Phase I Final Report, Human Effectiveness Directorate Biodynamics and Protection Division, Flight Motion Effects Branch, Brooks AFT, TX.

Elsmore, T., Reeves, D. (2002). *ARES™ User’s Manual*. Activity Research Services, Chula Vista, CA.

Girden, E. R. (1992). “ANOVA Repeated Measures,*” Sage University Publications Series on Quantitative Applications in the Social Sciences,* 7(84), 1-77.

Hursh, S. R. (2001). *Description of SAFTE™ Model, version 2*, Science Applications International Corporation, Joppa, MD.

Hursh, S. R., Redmond, D. P., Johnson, M. L., Thorne, D. R., Belenky, G., Balkin, T. J., Storm, W. F., Miller, J. C., & Eddy, D. R. (2004). “Fatigue Models for Applied Research in Warfighting,” *Aviation, Space, and Environmental Medicine,* 75(3), section II, Suppl., A44-A52.

Kronauer, R. E., & Stone, B. M. (2004). “Commentary on Fatigue Models for Applied Research in Warfighting,” *Aviation Space Environmental Medicine,* 75(3), section II, Suppl., A54-A56.

Kerkhof, F. A., & Van Dongen, H. P. (1996). ”Morning-type Individuals Differ in the Phase Position of their Endogenous Circadian Oscillator,” *Neuroscience Letters,* 218(3), 153-6.

LeDue, P. A. Jr., Caldwell, J. A. Jr., & Ruyak, P. S. (2000). “The Effects of Exercise as a Countermeasure for Fatigue in Sleep-deprived Aviators,” *Military Psychology*, 12(4), 249-266.

Mass, J. B., Wherry, M. L., Hogan, B., & Axelrod, D. J. (1998). *Power Sleep: The Revolutionary Program That Prepares Your Mind for Peak Performance,* New York, NY: Villard.

Matsumoto, Y., Mishima, K., Satoh, K., Shimizu, T., & Hishikawa, Y. (2002). “Physical Activity Increases the Dissociation between Subjective Sleepiness and Objective Performance Levels during extended Wakefulness in Humans,” *Neuroscience Letters*, 326(2), 133.

National Sleep Foundation, (n.d.). *2001 Sleep in America Poll*, Retrieved April 16, 2004, from http://www.sleepfoundation.org/publications/2001poll.cfm#4

Neri, D. F. (2004). “Preface: Fatigue and Performance Modeling Workshop, June 13-14, 2002,” *Aviation, Space, and Environmental Medicine*, 75(3), section II, Suppl., A1-A3.

Neri[1], D. F., Dinges[2], D. F., & Rosekind[1], M. R. (1997). *Sustained Carrier Operations: Sleep Loss, Performance, and Fatigue Measures.* [1]Fatigue Countermeasures Program Flight Management and Human Factors Division, NASA Ames Research Center, Moffet Field, CA, [2]Unit for Experimental Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, PA.

Newton, R. R., and Rudestam, K. E. (1999). *Your Statistical Consultant: Answers to your Data Analysis Questions,* Thousand Oaks, CA: Sage Publications, Inc.

O’Connor, M. J., & Patillo, P. L. “Reengineering Human Performance and Fatigue Research through Use of Physiological Monitoring Devices, Web-based and Mobile Device Data Collection Methods, and Integrated Data Storage Techniques,” Master’s Thesis, Naval Postgraduate School, Monterey, CA, December 2003.

Reeves, D., Winter, K., Kane, R., Elsmore, T., & Bleiberg, J. (2002). *ANAM™ 2001 User's Manual: Clinical & Research Modules.* National Cognitive Recovery Foundation, Special Report NCRF-SR-2002-1, Draft: February 2002.

Rosekind, M. R., Philippa, G. H., Gregory, K. B., Smith, R. M., Miller, D. L., Oyung, R., Webbon, L. L., & Johnson, J. M. (1996). "Managing Fatigue in Operational Settings 1: Physiological Considerations and Countermeasures," *Behavioral Medicine*, 21, 157-170.

Sagaspe, P. P., Taillard J., Moore, N., Guilleminault, C., Sanchez-Ortuno, M., Akerstedt, T., & Bioulac, B. (2003). "Fatigue, Sleep Restriction, and Performance in Automobile Drivers: a Controlled Study in a Natural Environment," *Sleep*, 26(3), 277-280.

Tassi, P., & Muzet, A. (2000). "Sleep Inertia," *Sleep Medical Review*, 4(4), 341-353.

Van Dongen, H. P. A. (2004). "Comparison of Model Predictions to experimental data: Rectifying False Impressions," *Aviation, Space, and Environmental Medicine*, 75(3), section II, Suppl., A122-A124.

Van Dongen, H. P. A., & Dinges, D. F. "Circadian Rhythms in Fatigue, Alertness and Performance", In: Kryger, T., Roth, T., Dement, C.W. (Eds.), *Principles and Practice of Sleep Medicine*, 3rd ed., 391-399, W.B. Saunders, Philadelphia, PA, 2000.

Van Dongen, H. P. A., Maislin, G., & Dinges, D. F. (2004). "Dealing with Inter-individual Differences in the Temporal Dynamics of Fatigue and Performance: Importance and Techniques," *Aviation, Space, and Environmental Medicine*, 75(3), section II, Suppl., A147-A154.

Van Dongen, H. P. A., Rogers, N.L., & Dinges, D. F. (2003). "Sleep debt: Theoretical and Empirical Issues", *Sleep and Biological Rhythms*, 1, 5-13.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Nita L. Miller
   Naval Postgraduate School
   Monterey, California

4. Laura A. Barton
   Naval Postgraduate School
   Monterey, California

5. Sam Buttrey
   Naval Postgraduate School
   Monterey, California

7. Tim Elsmore
   Activity Research Services
   Chula Vista, California

8. Steve Hursh
   Science Applications International Corporation
   Abingdon, Maryland

9. Dennis Reeves
   San Diego, California